

Levels of Processing Versus Transfer Appropriate Processing

C. DONALD MORRIS, JOHN D. BRANSFORD, AND JEFFERY J. FRANKS

Vanderbilt University

Levels of processing were manipulated as a function of acquisition task and type of recognition test in three experiments. Experiment 1 showed that semantic acquisition was superior to rhyme acquisition given a standard recognition test, whereas rhyme acquisition was superior to semantic acquisition given a rhyming recognition test. The former finding supports, while the latter finding contradicts, the levels of processing claim that depth of processing leads to stronger memory traces. Experiment 2 replicated these findings using both immediate and delayed recognition tests. Experiment 3 indicated that these effects were not dependent upon the number of times a rhyme sound was presented during acquisition. Results are interpreted in terms of an alternate framework involving transfer appropriate processing.

According to the levels of processing framework proposed by Craik and Lockhart (1972), the nature and duration of the memory trace is determined by the level or depth at which the input is processed. Inputs that receive only superficial analyses such as those prompted by nonsemantic orienting tasks are assumed to be more poorly retained than inputs subjected to deeper semantic analyses.

A large number of studies appear to support the levels of processing framework (e.g., Hyde & Jenkins, 1969; Till & Jenkins, 1973; Walsh & Jenkins, 1973). A somewhat smaller set of studies has pointed toward the need to further differentiate levels of processing within the semantic level of analysis (e.g., Craik & Tulving, 1975; Schulman, 1974; Klein & Saltz, 1976; Seamon & Murray, 1976) and to consider additional memory variables

such as retrieval constraints, trace uniqueness, trace congruity, and so forth (e.g., Craik & Tulving, 1975; Moscovitch & Craik, 1976). To our knowledge, however, no theorists have explicitly addressed the question of what is meant by superficial or nonmeaningful processing in contrast to semantic processing. We shall argue that particular acquisition activities are never inherently "superficial" or "nonmeaningful." Instead, task meaningfulness must be defined relative to particular learning goals.

Consider a recent experiment conducted within the levels of processing framework. Seamon and Murray (1976) presented orienting instructions indicating that subjects should either attend to word meaning (Task A) or attend to the position of their lips during vocalization of each input (Task B). Results indicated that the deeper processing suggested by Task A produced better memory than did the superficial, nonmeaningful processing prompted by Task B. Seamon and Murray also manipulated the normative meaningfulness of input stimuli and found that meaningfulness facilitated recall only for acquisition Task A.

It is instructive to ask why attention to the

This research was supported by Grant No. NEG-00-3-0026 from the National Institute of Education, awarded to the second and third authors. We gratefully acknowledge the assistance of Douglas Hintzman and an anonymous reviewer in providing especially helpful comments on an earlier draft of this paper. Reprints may be obtained from Donald Morris, Department of Psychology, Vanderbilt University, Nashville, Tennessee 37240.

position of one's lips during pronunciation constitutes a nonmeaningful or superficial level of processing. At first glance, the answer to such a question seems obvious. Subjects are not prompted to process the *meaning* of each word. However, is the failure to process the semantic meaning of inputs necessarily equivalent to performing a superficial or nonmeaningful task?

Assume that one wants to teach principles of speech perception and articulation to students. The present authors' experiences (as well as those of colleagues who have taught speech perception) suggest that an especially helpful teaching technique involves asking students to attend to the position of their lips and tongue while pronouncing words. Students usually find this to be an extremely *meaningful* exercise, despite the fact that they are not prompted to process the *semantic meaning* of the words used in the exercise. Indeed, the semantic meaning of the words presented is not necessarily a meaningful component of such an instructional exercise. Similarly, if one wants to teach students about rhyming, the semantic meaning of the words presented as illustrations is not necessarily a meaningful aspect of the task at hand.

Imagine that an instructor wishes to test the degree to which students learned from the above-mentioned classroom exercises. It would seem totally inappropriate to test students by asking them to identify the particular inputs (e.g., words) that were used during the learning exercises. The purpose of the speech perception and rhyme exercises is not to learn the particular inputs used as illustrations during instruction. Instead, the purpose is to learn sound-mouth and tongue relationships (for speech perception and articulation), to learn sound-sound relationships (for rhyming), and so forth. Such learning tasks are not necessarily shallow, superficial, or nonmeaningful. Do such tasks necessarily result in memory traces that are less adequate and durable than do other tasks which prompt subjects to process the semantic

aspects of inputs? We suggest that the answer to this question has not yet been subjected to appropriate experimental test.

Investigators who have utilized the levels of processing framework have been equivalent to a speech perception or rhyme instructor who tests his student on the particular words used to exemplify the desired to-be-learned information. In a classroom situation, it seems clear that sophisticated students would strongly object to such a testing procedure. Clearly, the tests would not be designed to tap what was supposed to be learned. The levels of processing claim that less meaningful or more superficial analyses of inputs result in less durable memory traces is therefore questionable or at least highly ambiguous. It seems clear that one can test people inappropriately and therefore find evidence for "inadequate processing or learning" relative to that particular testing criterion. Will the superiority of deep, semantic processing persist irrespective of the test one employs? For example, assume that one group of subjects is asked to check for *es* in each word (Hyde & Jenkins, 1969), whereas a second group is prompted to process each word at a deeper, semantic level. Which group will be better and faster at estimating the number of acquisition stimuli that contained an "*e*?"

The present experiments are designed to begin an initial inquiry into the assumptions underlying the levels of processing framework. Their purpose is to explore the degree to which assumptions about the "goodness" of particular acquisition activities must be defined relative to particular learning goals (as well as to tests designed to be congruent with these goals). It is possible that so-called superficial or nonmeaningful tasks (e.g., rhyme acquisition) are inferior to semantic tasks irrespective of the testing situation. Alternatively, different modes or levels of processing may simply allow people to acquire different sorts of information, each of which may have the potential for being equally strong and durable (as revealed by appropriate

testing situations). The present experiments provide an initial investigation of these possibilities.

Two incidental learning tasks were given to subjects to induce them to process verbal items at different levels. One task required subjects to judge the appropriateness of a target word within the context of a given sentence frame. The second task involved judging whether or not a target word rhymed with another word. The first task presumably tapped a semantic (deep) level of processing, whereas the second task induced subjects to process the words at a phonetic (a relatively more superficial) level. The levels of processing framework predicts that the semantic processing task should result in superior retention of items, relative to the phonetic processing task. But, as argued in the introduction, tests of this claim have generally used targets and foils that may be considered appropriate to and dependent upon semantic modes of processing. Thus, many of the results favoring the levels of processing claims may be due in large part to an inherent bias in the way in which memory was tested.

To provide a more adequate test of these claims it is necessary to consider not only the processing levels dictated by the acquisition tasks, but also the processing levels induced by the form of the memory test itself. Accordingly, the present experiments factorially varied the kind of acquisition task with the kind of memory test given to subjects. It was expected that these experimental manipulations would directly affect the levels at which subjects processed materials during acquisition and testing.

The basic paradigm of the present experiments is as follows: Subjects were given either a semantic or a phonetic orienting task. Half of the subjects in each experiment were given a recognition test in which targets were the *original* items presented during acquisition. The other half of the subjects received a recognition test in which targets were *rhymes* of the items presented originally. Thus, the

level of processing for a set of items was varied not only at acquisition, but at the time of test as well.

EXPERIMENT 1

Experiment 1 was designed to investigate the possibility of an interaction between acquisition mode and type of memory test, given an immediate testing situation. It was hypothesized that a semantic level of processing during acquisition would indeed facilitate recognition performance relative to a rhyme-related level of processing, given a standard recognition test. This superiority was expected to be neutralized, and possibly reversed, however, when subjects were given a rhyming recognition test.

Method

Subjects. Thirty-two subjects from an introductory psychology class served in this experiment. Each subject was run individually.

Design. A $2 \times 2 \times 2$ factorial design was utilized. There were two types of Acquisition Tasks, a Semantic orienting task and a Rhyme orienting task. These tasks induced deep and superficial levels of processing, respectively. Nested within this factor was a Congruency factor (see Craik & Tulving, 1975). Target words were either congruent or not with a particular context for both kinds of Acquisition Task. For example, targets were either meaningful or nonmeaningful within a particular sentence, or targets either rhymed or did not rhyme with the last word of the preceding sentence. The two levels of this factor are referred to as Yes and No, respectively. Both Acquisition Task and Congruency were varied within subjects. The third factor, Type of Test, was manipulated between subjects. Sixteen subjects were given a Standard recognition test in which target items were the *original* acquisition items. The other sixteen subjects were given a Rhyming recognition test in which targets were *rhymes* of the original items.

Materials and procedures. The experimental stimuli were 32 target words, embedded in sentences. These words were subsequently presented on a recognition test along with 32 foils. All targets and foils were common five-letter words, of either one or two syllables. Each word was chosen to meet the following constraints: (1) It bore little semantic or phonetic similarity to any other word on the list; and (2) at least one other word rhymed with it if it was a foil, or at least two other words rhymed with it if it was a target. (The reasoning for this last constraint is given below.)

During acquisition, the experimenter read aloud 32 sentences with one word (the target) missing from each. The word "BLANK" was said in place of the missing target word. Each sentence was followed by a 2-second pause, then the vocal presentation of the target word. There were four types of sentences, each with eight instances, and each representing a particular mode of acquisition. These acquisition modes corresponded to the four possible combinations of the within-subject factors: Semantic-Yes, Semantic-No, Rhyme-Yes, and Rhyme-No. Sentences requiring semantic processing of the targets were of the form "The _____ had a silver engine." Sentences inducing phonetic processing of the targets were of the form "_____ rhymes with legal." The target word was either a Yes or No. For example, given "The _____ had a silver engine," presentation of the target "TRAIN" would represent a Semantic-Yes acquisition mode, whereas presentation of the target "EAGLE" would represent a Semantic-No acquisition mode. Analogously, for a sentence such as "_____ rhymes with legal," presentation of targets such as "EAGLE" and "PEACH" represent Rhyme-Yes and Rhyme-No acquisition modes, respectively. Each subject received the same 32 target words. These words appeared in one of four different random orders and each target was presented equally often as an instance of each of the

four acquisition modes. These last two considerations were both counterbalanced across subjects.

All sentences were short and constructed in such a way that subjects could easily decide if the target was meaningful or nonmeaningful, or a rhyme or nonrhyme. Subjects were simply to respond positively or negatively as to the appropriateness of each target given the immediately preceding sentence. Subjects were told that their responses were to serve as normative data for future experiments.

After receiving the 32 acquisition sentences, subjects were immediately given a recognition test consisting of 64 items, 32 targets and 32 foils in a random order. These words were read aloud by the experimenter at a rate of one word approximately every 5 seconds. Subjects responded on a binary yes-no scale and then responded on a 5-point scale rating their confidence of their recognition responses. For those subjects receiving the Standard recognition test, the to-be-remembered items were those they had received during acquisition. For those receiving a Rhymingrecognition test, to-be-remembered items were words that rhymed with the original acquisition items. As mentioned before, all foils (e.g., "POUND") had at least one rhyme (e.g., "MOUND"). It was necessary, however, to choose targets (e.g., "EAGLE") that had at least two rhymes, one given in the acquisition sentences (e.g., "LEGAL") and one given in the recognition test (e.g., "REGAL").

Results

Table 1 presents a summary of the mean corrected recognition scores for each experimental condition. It was considered inappropriate to use uncorrected scores, since these would not take into account the response biases of individual subjects, that is, their inclinations to respond positively or negatively regardless of the acquisition mode of a particular item. In an attempt to remove at least part of this potential response bias, false positive responses were subtracted from the

TABLE 1

MEAN CORRECTED PROPORTION SCORES AS A
FUNCTION OF TYPE OF RECOGNITION TEST AND
TYPE OF ACQUISITION MODE

Acquisition mode	Recognition test	
	Standard	Rhyming
Semantic-Yes	.844 (.155) ^a	.333 (.224)
Rhyme-Yes	.633 (.239)	.489 (.252)
Semantic-No	.859 (.163)	.325 (.236)
Rhyme-No	.524 (.271)	.184 (.166)

^a Numbers in parentheses represent standard deviations.

scores in each condition. The corrected scores are presented as proportions and were obtained in the following manner: Suppose a subject had correctly recognized six Semantic-Yes items and four Rhyme-Yes items but had also responded positively to eight of the 32 foils. The corrected scores would then be $(6/8) - (8/32) = 0.50$ and $(4/8) - (8/32) = 0.25$, respectively.

A survey of these means given that subjects received a Standard recognition test reveals that the Semantic acquisition mode was indeed superior to the Rhyme acquisition mode. This finding supports the levels of processing claim that greater depth of processing leads to improved memory traces (e.g., Craik & Tulving, 1975). An examination of the Rhyming recognition scores suggests quite a different account, however. In this condition, memory performance was better when subjects had been given a Rhyme-Yes acquisition mode relative to a Semantic-Yes acquisition mode (although this did not hold for No acquisition sentences, see below). This latter effect is difficult to account for within the levels of processing framework.

A $2 \times 2 \times 2$ (Type of Test \times Acquisition Task \times Congruency) analysis of variance on the corrected proportion scores showed a significant main effect for Type of Test, $F(1, 30) = 39.09, p < .001$, indicating a general

superiority of the Standard recognition test over the Rhyming recognition test; a significant main effect for Acquisition Task, $F(1, 30) = 21.31, p < .001$, indicating a general superiority of the Semantic acquisition mode over the Rhyme acquisition mode; and a significant Type of Test \times Acquisition Task interaction, $F(1, 30) = 23.89, p < .001$, discussed more fully below. The analysis of variance also showed a significant Congruency effect, $F(1, 30) = 9.87, p < .004$, and a significant Acquisition Task \times Congruency interaction, $F(1, 30) = 15.91, p < .001$. In general, targets embedded within congruent sentence frames were better remembered than those embedded in incongruent frames, replicating the findings of Craik and Tulving (1975). This effect was primarily attributable to the enhanced recognition of Rhyme-Yes over Rhyme-No scores.

The effect of primary interest in these scores is the observed interaction between Acquisition Task and Type of Test. This interaction was obtained in the analysis of variance with scores collapsed across the Yes-No acquisition dimensions. An examination of Table 1 indicates that the nature of this interaction varies substantially between the Yes and No conditions. Inspection of the upper half of Table 1 (for Yes acquisition conditions) reveals that, given a Standard recognition test, the Semantic acquisition mode was superior to the Rhyme acquisition mode, $.844 > .633, t(30) = 3.99, p < .001$. However, given that subjects received a Rhyming recognition test, the Rhyme acquisition mode provides superior recognition performance relative to the Semantic acquisition mode, $.489 > .333, t(30) = 2.95, p < .01$. If we now consider the lower half of Table 1 (for No acquisition conditions), it is evident that the Semantic acquisition mode was superior to the Rhyme acquisition mode for both types of recognition test, although the superiority was greater for the Standard test. Since there was no reversal here, as there was for Yes scores, no further statistical tests were considered necessary.

One reason for these results with No items may be the lack of integration of the target and its sentence frame for No acquisition sentences. This would lead to a decrement in performance, particularly for the Rhyme-No conditions. Craik and Tulving (1975), for instance, have noted that the role of congruency in encoding and retrieval operations needs further clarification. Its influence, particularly with other levels of processing factors, remains unclear at present.

A second reason why Rhyme-No conditions may have poor performance on the Rhyming test stems from the following consideration. Rhyme-No acquisition frames are of the form "_____ rhymes with ditch: LEGAL." Under these conditions LEGAL is assumed to be the target item and the rhyme transfer test assesses a subject's abilities to detect a rhyme word like REGAL. Although this procedure was utilized in the present studies in order to be congruent with previous investigations (e.g., Craik & Tulving, 1975; Moscovitch & Craik, 1976), there are certain reasons to doubt its adequacy. The focal point of a sentence like "_____ rhymes with ditch: LEGAL" may very well be "ditch" and not LEGAL. If subjects could somehow be instructed to understand that words rhyming with "ditch" were the target items for Rhyme-No sentences and were later tested with words like WITCH (rather than REGAL), performance might greatly improve on the Rhyming test, despite the lack of a sentence-target congruency effect. This possibility must await further research. In the meantime, it seems more appropriate to focus present analyses and discussions on the Yes items rather than the No items.

An analysis of the obtained confidence ratings provided no additional insights into the nature of the effects reported above.

EXPERIMENT 2

Experiment 2 was designed with two functions in mind. One function was to

replicate the findings of Experiment 1. To this end, certain changes were made to provide greater reliability and power for the statistical tests of the phenomenon. The result to be replicated indicates that memory performance seems to be affected by processing level during acquisition task *and* by processing level during retention test. Given this finding, the levels of processing claim that semantic orienting tasks provide inherently stronger memory traces stands in need of qualification. The levels of processing framework also claims that semantic processing results in longer-lasting or more durable memory traces. Thus, the superiority of "semantic" traces over "rhyme" traces should increase, or at least remain constant, over time. The second function of Experiment 2 was to test this prediction, using the same basic paradigm as before. In this experiment, half the subjects received the recognition test immediately, as before, while the other half were tested after a 24-hour delay. Thus, it was hoped that more could be ascertained regarding the interaction between levels of processing at acquisition and test.

Method

Subjects. One hundred and fourteen subjects from an introductory psychology class served in this experiment. Subjects were run in 16 groups of six to eight persons each.

Design. The design was similar to that of Experiment 1 with the exception of an additional between-subjects factor, Time of Test. There were two levels of this factor, Immediate and Delayed. Thus, the design was now a $2 \times 2 \times 2 \times 2$ factorial design.

Materials and procedures. The materials were identical to those of Experiment 1 with the following exception: In Experiment 1 there were 16 distinct stimulus lists (four random orders \times four acquisition modes per target), whereas the present experiment utilized only eight distinct stimulus lists (two random orders \times four acquisition modes per target).

There were two procedural variations from

TABLE 2

MEAN CORRECTED PROPORTION SCORES AS A FUNCTION OF TYPE OF TEST, TIME OF TEST, AND TYPE OF ACQUISITION MODE

Acquisition mode	Immediate		Delayed	
	Standard	Rhyming	Standard	Rhyming
Semantic-Yes	.757 (.169) ^a	.300 (.151)	.450 (.208)	.180 (.180)
Rhyme-Yes	.682 (.180)	.424 (.220)	.418 (.188)	.291 (.229)
Semantic-No	.699 (.206)	.299 (.176)	.387 (.209)	.238 (.221)
Rhyme-No	.528 (.192)	.272 (.251)	.249 (.171)	.166 (.218)

^a Numbers in parentheses represent standard deviations.

Experiment 1: (1) Subjects could now respond "yes," "no," or "unsure" after each acquisition sentence, and (2) eight groups of subjects were tested immediately, whereas eight were tested after a 24-hour delay. To help insure the incidental nature of the task, subjects were told that they were to provide additional normative ratings on the following day and that the experiment concerned the reliability of such ratings over time.

Results

Table 2 presents a summary of mean corrected proportion scores for each experimental condition. A survey of these means reveals again that a Semantic acquisition mode is superior to a Rhyme acquisition mode when subjects were given a Standard recognition test. Thus, this finding replicates that found in Experiment 1 and supports the levels of processing claim. In addition, this superiority persists when subjects are tested after a 24-hour delay, although the absolute difference between Semantic and Rhyme acquisitions has diminished. When subjects are given a Rhyming recognition test, a different pattern of results appears. As in Experiment 1, memory performance was better when subjects had been given a Rhyme-Yes acquisition rather than a Semantic-Yes acquisition. This effect holds not only for the Immediate condition (thus replicating the results of Experiment 1) but for the Delayed

condition as well. It is interesting to note that the absolute difference between Semantic-Yes and Rhyme-Yes acquisitions remains fairly constant over the delay period. Note once again, however, that no reversal was found for the No scores. The Semantic-No condition was superior to the Rhyme-No condition for both the Standard and the Rhyming tests and for both the Immediate and the Delayed conditions.

A $2 \times 2 \times 2 \times 2$ analysis of variance (Type of Test \times Acquisition Task \times Congruency \times Time of Test) on the corrected proportion scores showed significant main effects for Type of Test, $F(1, 110) = 86.51, p < .001$; Acquisition Task, $F(1, 110) = 5.37, p < .02$; Congruency, $F(1, 110) = 30.03, p < .001$; and Time of Test, $F(1, 110) = 54.18, p < .001$. These represent, in general, the superiority of Standard over Rhyming recognition tests, Semantic over Rhyme acquisition modes, Yes over No acquisition modes, and Immediate over Delayed tests, respectively. Significant interactions were found for Type of Test \times Time of Test, $F(1, 110) = 11.88, p < .001$; Type of Test \times Acquisition Task, $F(1, 110) = 21.24, p < .001$; and Congruency \times Acquisition Task, $F(1, 110) = 16.71, p < .001$. The Type of Test \times Time of Test interaction was due to a larger difference between the Standard and Rhyming recognition tests when tested immediately than when tested after a 24-hour delay. The Congruency \times Acquisition Task

interaction was primarily attributable to the relatively poor performance levels of the Rhyme-No acquisition mode relative to the other three acquisition modes. The Type of Test \times Acquisition Task is described below.

For the reasons expressed in the discussion of the results of Experiment 1, it is difficult to interpret differences between Yes and No acquisition modes. Therefore, the critical Type of Test \times Acquisition Task interaction was examined with respect to only the Yes scores. When subjects were given a Standard recognition test, Semantic acquisition was better than Rhyme acquisition for the Immediate test condition, $.757 > .682$, $t(110) = 1.66$, $p < .10$. Although this difference failed to reach conventional levels of significance for a two-tailed test, it was significant for a one-tailed test, $p < .05$. For the Delayed test condition, Semantic acquisition was also better than Rhyme acquisition, but again the advantage was not significant, $.450 > .418$, $t < 1.00$. Though these comparisons fail to suggest significant differences, the differences are in the direction predicted by the levels of processing claim. When subjects were given a Rhyming recognition test, Rhyme acquisition was significantly better than Semantic acquisition for both the Immediate and the Delayed tests, $.424 > .300$, $t(110) = 2.66$, $p < .01$ and $.290 > .180$, $t(110) = 2.38$, $p < .02$, respectively. Thus, the levels of processing claim that greater depth of processing results in stronger memory traces is, again, not supported.

The second claim, that depth of processing also provides more durable traces, is also questioned by the present data. The present authors had assumed that rhyme processing might yield extremely poor performance on a rhyme test after a 24-hour delay. In contrast, semantic processing might well show much less decrement in performance on the delayed rhyme test. However, the three-way interaction among Time of Test, Type of Test, and Acquisition Task was not significant. Furthermore, an examination of the Yes scores in Table 2 reveals that the failure of the three-way

interaction to reach significance was not due to a lack of statistical resolution. Instead, it appears that the results run in a direction opposite to that discussed above. For instance, given a Standard recognition test, the superiority of Semantic over Rhyme acquisition diminishes over time ($.757$ v $.682$ in the Immediate condition compared to $.450$ v $.418$ in the Delayed condition). This in itself is not surprising, but it offers an interesting contrast to the Rhyming recognition condition. In this condition, the superiority of Rhyme over Semantic acquisition diminishes to a lesser extent over time ($.424$ v $.300$ in the Immediate condition compared to $.290$ v $.180$ in the Delayed condition). A comparison of the differences between differences ($.075 - .032 = .043$ v $.124 - .110 = .014$, respectively) failed to reach statistical significance. Thus, these results question the assumption that semantic processing provides more durable memory traces.

Finally, there are two major questions regarding the recognition performance of Rhyme-Yes words. Examination of Tables 1 and 2 shows that Rhyme-Yes acquisition stimuli are better retained on Standard, rather than Rhyming, recognition tests. Although the present interest concerns the Acquisition Task \times Type of Test interaction, these findings warrant further consideration. Two possibilities suggest themselves. One, the Standard recognition test is simply an easier test to perform. Two, and perhaps more importantly, items on the Standard recognition test contain both semantic and phonetic information relevant to correct identification of the target. Items on the Rhyming recognition test contain rhyme information which relates only to phonetically relevant information but not to semantic information. This possibility is currently being researched.

The second point to be considered involves an inherent procedural difference between Rhyme-Yes and Semantic-Yes conditions. In particular, the Rhyme-Yes condition includes two occurrences of items that rhyme with the

Rhyming test target items (e.g., "——— rhymes with legal: EAGLE"), whereas the Semantic-Yes condition includes only one occurrence of an item that rhymes with the Rhyming test target items (e.g., "A —— has feathers: EAGLE"). Clearly, it takes two items to define a rhyme. However, it is possible that this procedural difference nevertheless represents a confound in the present design.

The procedural differences between semantic and rhyme modes of acquisition are important and in need of further consideration. For example, Rhyme-Yes acquisition involves two sources of information that might be utilized in subsequent transfer situations (e.g., a Rhyming test); Semantic-Yes acquisition involves only one. One way to control for this difference is to present two sources of rhyme-related information for Semantic-Yes items as well. Thus, given a Rhyming recognition test, Semantic-Yes acquisition would not be at a disadvantage (in numerical terms) relative to Rhyme-Yes acquisition. Experiment 3 controls for this procedural factor in order to examine its role in determining previous results.

EXPERIMENT 3

Method

Subjects. Twenty-five subjects from an introductory psychology class served in this experiment. Subjects were tested in four groups of six or seven persons each.

Design. A 2×2 factorial design was utilized. The two factors were Acquisition Task and Congruency, as in Experiments 1 and 2. All subjects were given a Rhyming recognition test.

Materials and procedures. The materials were identical to those of Experiments 1 and 2 with the following exceptions: Only four distinct stimulus lists were used; the acquisition mode of each target was varied across these lists. A second change concerns the construction of the semantic encoding sentences. These sentences were altered so that

four Semantic-Yes sentences and four Semantic-No sentences would have as their last word a rhyme of one of the eight Semantic-Yes acquisition target items. This rhyme word was not the Rhyming recognition test target nor did the rhyme word ever appear in the same sentence as the corresponding acquisition target. Consider the Semantic-Yes item EAGLE, which might appear in the following context "The —— has feathers: EAGLE." Now, a rhyme of EAGLE, such as "regal," would appear at the end of another Semantic sentence frame. For example, subjects might hear "The —— was very regal: LOCAL." The interval between presentation of a Semantic-Yes target (e.g., EAGLE) and its rhyme word (e.g., "regal") was variable.

The nature of the question under investigation concerns whether or not Rhyme-Yes items have an unfair advantage over Semantic-Yes items on a subsequent Rhyming recognition test. For this reason it was considered unnecessary to test subjects on a Standard recognition test. Therefore, Type of Test was not a factor in this experiment; all subjects received a Rhyming test.

Results

Table 3 presents a summary of the mean corrected proportion scores for each experimental condition. These results are in general

TABLE 3

MEAN CORRECTED PROPORTION
SCORES FOR RHYMING RECOGNITION
AS A FUNCTION OF ACQUISITION
MODE

Acquisition mode	Rhyming test
Semantic-Yes	.303 (.203) ^a
Rhyme-Yes	.446 (.187)
Semantic-No	.283 (.158)
Rhyme-No	.265 (.174)

^a Numbers in parentheses represent standard deviations.

agreement with those of Experiments 1 and 2. Of particular importance is the superiority of Rhyme-Yes acquisition over Semantic-Yes acquisition. This superiority holds even when Semantic-Yes items have a second presentation of the critical sound information during acquisition.

A 2×2 (Acquisition Task \times Congruency) analysis of variance on the corrected proportion scores showed a significant main effect for Acquisition Task, $F(1, 24) = 6.09$, $p < .02$; a significant main effect for Congruency, $F(1, 24) = 13.10$, $p < .002$; and a significant Acquisition Task \times Congruency interaction, $F(1, 24) = 10.84$, $p < .003$. These effects are generally attributable to the enhanced recognition for Rhyme-Yes items. The critical comparison of this experiment is between Rhyme-Yes and Semantic-Yes recognition scores. This comparison yields highly significant results, $.446 > .303$, respectively; $t(24) = 4.15$, $p < .001$. Thus, it appears that Rhyme-Yes acquisition is superior to Semantic-Yes acquisition when subjects are tested on Rhyming recognition. This effect persists even when Semantic-Yes items have had a second presentation within the acquisition list of the rhyme-related information relevant for a Rhyming test.

OVERALL SUMMARY AND CONCLUSIONS

The results of the present studies suggest a need to reconsider certain assumptions basic to the levels of processing framework. In particular, arguments that nonsemantic or shallow levels of processing are necessarily inferior to deeper levels of processing are questionable. To be sure, so-called nonsemantic levels of processing may look inferior if subjects are asked to perform subsequent test tasks (e.g., reproducing the nominal stimuli) that are not directly related to what was learned during acquisition. In such cases, however, the reasons for the inferiority may be due to the inappropriateness of the relationship between acquisition and test rather

than the inherent inferiority of the acquired memory traces. If one wants to know the number of words in a list that contains *es*, the number containing two syllables, and so forth, then "*e*-checking" or "syllable counting" would appear to be appropriate modes of processing. Similarly, if one wants to learn about rhyming information, it would seem beneficial to pay attention to the rhymes of words. In the present studies, acquisition manipulations that directed subjects to attend to the rhymes of inputs resulted in better performance on a rhyming test than did acquisition activities that prompted subjects to process the "semantic meaning" of inputs. Similar results were obtained following a 24-hour delay between acquisition and test and when number of potential rhyme items was controlled.

Results such as these suggest that it might be useful to replace the concept of "levels of processing" with one emphasizing "transfer appropriate processing." The latter concept emphasizes that the value of particular acquisition activities must be defined relative to particular goals and purposes. Furthermore, assumptions about the quality and durability of the resulting memory traces can only be determined relative to the appropriateness of the testing situation. The concept of transfer appropriate processing suggests that it is no longer beneficial to simply assume that the traces of certain items are less durable or adequate than others because those items were processed at a shallower level. The evidence that appears to support this latter assumption involves test situations that are not optimal for assessing what was actually learned.

The current assumptions about the potential durability of even superficial memory traces are congruent with results from other studies. For example, a number of researchers report data indicating that so-called superficial aspects of encoding activities (e.g., orthographic case, sound of voice, orientations of presented words or sentences) are retained for surprisingly long periods of times (see

Arbuckle & Katz, 1976; Craik & Kirsner, 1974; Kirsner, 1973; Kolers, 1975, a, b; Kolers & Ostry, 1974). Note that these latter studies did not prompt subjects to orient solely to these superficial aspects at time of acquisition and then test them on memory for the particular inputs exemplifying these aspects. Instead, subjects apparently processed the inputs at a number of levels (including the semantic level). The data indicate that superficial as well as semantic information played a role in remembering, thereby indicating that the former information was in fact processed and must indeed be capable of being well retained.

One further aspect of the present data needs to be considered. In the present experiments, rhyme acquisition was superior to semantic acquisition for the rhyme test. Overall, however, the semantic acquisition-semantic test conditions resulted in better performance than the rhyme acquisition-rhyme test conditions. Do such results necessitate a concession to the levels of processing claim that superficial nonsemantic processing results in inherently less adequate and durable memory traces? In actuality, the present results suggest that so-called superficial aspects of "memory traces" were at least as durable as semantic traces (i.e., both were maintained over 24 hours). Nevertheless, in terms of absolute values, there is a superiority of semantic acquisition-semantic test over nonsemantic acquisition-nonsemantic test given both immediate and delayed tests. Do data such as these thereby require acceptance of the levels of processing approach?

The beginnings of an alternate conceptualization of data such as those noted above have been discussed by Jenkins (1974). He asks whether optimal memory might be most fruitfully viewed as being a function of (a) semantic processing or (b) skills possessed by the learner-rememberer. An implication of his discussion on skills is that semantic modes of processing may result in better memory for most college students, not because of any

inherent advantages of semantic memory traces, but because college students are usually primed to utilize their semantic skills in an experimental setting. On the other hand, consider that an experienced poet or an expert in speech perception, linguistic dialects, and so forth may be as efficient at remembering certain types of auditory information (given appropriate testing situations) as she or he is at remembering semantically processed information. This approach suggests a need to formulate theoretical conceptualizations of memory that do not simply assume that certain types of memory traces are inherently inferior because of the "level" at which they were processed.

Following the lead of Jenkins (1974) let us assume that the adequacy and durability of memory traces are a function of whether or not a subject *has* and *uses* appropriate knowledge and skills to precisely comprehend (encode) each input and hence differentiate it from other potential inputs. In contrast to the levels of processing framework, let us further assume that there are no inherent differences in the nature of the memory traces resulting from semantic versus nonsemantic levels of processing. Instead the emphasis is on the activation of appropriate skills and knowledge structures that "set the stage" for knowing precisely how and *in what ways* certain inputs differ from other potential inputs (e.g., see Bransford & Franks, 1976). The importance of differentiation has been discussed elsewhere (e.g., Gibson, 1940; Saltz, 1971). We emphasize that *even semantic processing* may not facilitate remembering if it does not result in precise differentiation of the acquisition and test stimuli (e.g., see Bransford, McCarrell, Franks, & Nitsch, in press; Stein, in press). Furthermore, the term "semantic processing" is usually used in a very ambiguous way (see also Postman, 1975).

Note that the term "semantic processing" is not necessarily equivalent to meaningful processing and that "nonsemantic processing"

is not necessarily equivalent to meaningless or superficial processing. For example, we have suggested that the linguistic semantic meanings of words presented during an exercise designed to teach speech articulation or rhyming are not necessarily meaningful components of the task. Similarly, the semantic meaning of the first and last words in a sentence like *Altitude precedes window* are not relevant for understanding the sentence. Instead, comprehension is enhanced by focusing on orthography and noting that "A" precedes "W" on an alphabetical scale (cf. Bransford, Nitsch, & Franks, 1976). Studies of comprehension (e.g., Bransford & Johnson, 1973; Bransford & McCarrell, 1975; Bransford & Franks, 1976; Bransford, Nitsch, & Franks, 1976) suggest a need to distinguish between the *semantic meaning* of inputs and their *understood meaning* (i.e., their significance). Indeed, perceptual artifacts, gestures, brush strokes, sounds, and so forth may or may not be *meaningful* to a person depending on whether or not an appropriate knowledge framework, or set of skills, is both available and activated at the time (e.g., see Bransford & McCarrell, 1975; Bransford, Nitsch, & Franks, 1976).

The preceding discussion suggests that even superficial aspects of inputs can be meaningful, depending on the knowledge possessed by subjects. In order to be well remembered, subjects also need to be able to use their past knowledge and skills to comprehend (encode) inputs in precise and unique ways. It seems useful to note that the current assumptions about the use of past knowledge structures to set the stage for more precisely comprehending (encoding) the unique aspects of particular inputs are different from assumptions about uniqueness that have appeared in the levels of processing literature. For example, a recent article by Moscovitch and Craik (1976) argues that uniqueness is somewhat important, but only in addition to assumptions about levels of processing.

The latter authors cite studies by Goldstein

and Chance (1970) indicating that nonmeaningful visual patterns are poorly recognized even though each pattern is unique, as well as an experiment by Craik and Tulving (1975) indicating that recognition of case encoded words was not facilitated even though the number of such instances was reduced from 40 to four items.

From the present perspective, the Craik and Tulving (1975) experiment did not test what was actually learned. It would be more appropriate to test memory for orthographic case information than for the individual inputs illustrating such cases. Furthermore, the Goldstein and Chance (1970) study says nothing about the stage-setting role of past knowledge for helping one uniquely encode an input and hence precisely differentiate it from other inputs. Appropriate past knowledge permits one to know precisely how and in what ways particular inputs (plus acts of encoding them; e.g., see Kolers & Ostry, 1974) differ from other aspects of one's knowledge. Without appropriate knowledge structures, one lacks precision with respect to differentiation (see especially Garner, 1974). For example, Indonesian words would be unique for most English speaking subjects, but the latter subjects would have few knowledge structures for uniquely differentiating each word from other things known (especially from one another). Precise memory for such "unique" inputs would therefore be quite poor.

Note, however, that inputs (e.g., words from another language, nonsense syllables, etc.) can always be differentiated from the rest of one's knowledge at *some* level. They are never totally meaningless. Subjects hearing a list of nonsense syllables could easily differentiate them from meaningful words, pictures of triangles, and so forth in a forced choice recognition experiment. Furthermore, such abilities to differentiate would probably persist over a relatively long period of time. The importance of focusing on levels at which people are asked to differentiate inputs from

other things that they know is reflected in the large effects on recognition of the foil items. Even the memory traces of relatively non-meaningful unique items may seem quite durable if subjects are tested with foils that can be differentiated at the level of well-known, higher-order invariants that were perceived during acquisition. These more abstract, higher-order invariants seem to be more readily acquired and remembered than lower-level specifics, much as subjects in the Bransford and Franks (1971) experiments seem more apt to acquire higher-level semantic invariants than they are to acquire the precise acquisition inputs that were heard (see Bransford, McCarrell, Franks, & Nitsch, 1977). Arguments about "trace strength" or "durability" must therefore be defined relative to the levels of precision at which subjects are asked to, or able to, differentiate, as well as defined relative to the appropriateness of the test tasks.

In conclusion, it appears useful to summarize the differences between the present approach and the levels of processing framework, particularly since Craik and Tulving (1975) and Moscovitch and Craik (1976) (see also Lockhart, Craik, & Jacoby, 1975) have suggested modifications of the original formulation proposed by Craik and Lockhart (1972). In particular, these authors suggest that "spread of encoding" or "encoding elaboration" may also be important determinants of memory. At the same time, they maintain assumptions about inherent, qualitative differences among various processing levels or domains. For example, Craik and Tulving (1975) state:

We assume that "depth" still gives a useful account of the major qualitative shifts in a word's encoding (from an analysis of physical features through phonemic features to semantic properties). Within one encoding domain, however, "spread" or "number of encoded features" may be better descriptions. (p. 34)

Assumptions of inherent, qualitative differences between nonsemantic and semantic

levels of processing also seem to play important roles in theorizing about spread or elaboration of encoding. For example, Craik and Tulving (1975) state:

It should be borne in mind that retention depends critically on the qualitative nature of the encoding operations performed—a minimal semantic analysis is more beneficial for memory than an elaborate structural analysis. (p. 48)

The present data suggest that evidence for this latter assumption is based on inadequate tests of what was learned, that is, a "structural" analysis of inputs was more beneficial than a "semantic" analysis for a subsequent rhyme test.

It seems clear that the two above-mentioned quotes from Craik and Tulving (1975) refer to the nature of the "traces" necessary for remembering the actual inputs (e.g., words) presented during acquisition. However, the levels of processing approach uses the term memory trace in an ambiguous manner. There are important differences between use of the term memory trace to refer to a trace of the nominal stimulus and the use of the term trace to refer to the result of a particular learning experience.

The major difference between the levels of processing and transfer appropriate processing frameworks involves their orientations toward the general problem of learning. In the current literature, the term "learning" is usually used synonymously with "learning a list of inputs," and the test is usually a test of memory for these inputs (e.g., see Craik & Tulving, 1975). The value of particular acquisition or learning activities is assessed in relation to the goal of remembering the acquisition inputs. However, even given the goal of remembering inputs, assumptions about the value of particular types of acquisition activities must be defined relative to the type of activities to be performed at the time of test. For example, acquisition processes optimal for recognition are not necessarily optimal for free recall or cued recall, and vice versa (e.g., see Tversky, 1973; Tulving & Thomson, 1973; Bransford,

Nitsch, & Franks, 1976). Moreover, the transfer appropriate processing framework goes beyond measures of people's ability to remember the actual inputs presented during acquisition. The problem of learning is broader than this.

The present orientation assumes that learning involves learning from inputs as well as learning inputs (e.g., see Bransford & Franks, 1976; Bransford & Nitsch, in press; Hannigan, Note 1). For example, attention to the position of the lips and tongue can allow one to *learn from* a set of inputs presented as examples, but it will not necessarily allow one to learn (and hence remember) the exact inputs. Depending on what one desires people to learn (e.g., sound-mouth and tongue relationships; sound-sound relationships), processes optimal for learning may therefore be different from those optimal for *remembering* the exact inputs presented during acquisition. Tasks designed to help people learn about speech sounds, word syllables, orthography, and so forth are not necessarily nonmeaningful, and nonsemantic levels of processing need not result in inherently inferior traces representing what was learned. In short, transfer appropriate processing may sometimes involve the superficial levels of analysis that are deemed less adequate by the levels of processing approach.

REFERENCES

- ARBUCKLE, T. Y., & KATZ, W. A. Structure of memory traces following semantic and nonsemantic orientation tasks in incidental learning. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 362-369.
- BRANSFORD, J. D., & FRANKS, J. J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, **2**, 331-350.
- BRANSFORD, J. D., & FRANKS, J. J. Toward a framework for understanding learning. In G. H. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press, 1976.
- BRANSFORD, J. D., & JOHNSON, M. K. Consideration of some problems of comprehension. In W. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.
- BRANSFORD, J. D., & MCCARRELL, N. S. A sketch of a cognitive approach to comprehension: Some thoughts about understanding what it means to comprehend. In W. Weimer & D. Palermo (Eds.), *Cognition and the symbolic processes*. Hillsdale, New Jersey: L. Erlbaum Associates, 1975.
- BRANSFORD, J. D., MCCARRELL, N. S., FRANKS, J. J., & NITSCH, K. E. Toward unexplaining memory. In R. E. Shaw & J. D. Bransford (Eds.), *Perceiving, acting and knowing: Toward an ecological psychology*. Hillsdale, New Jersey: L. Erlbaum Associates, 1977.
- BRANSFORD, J. D., & NITSCH, K. E. Coming to understand things we could not previously understand. In J. F. Kavanagh & W. P. Strange (Eds.), *Implications of basic speech and language research for the schools and clinic* (working title). Cambridge, Massachusetts: MIT Press, in press.
- BRANSFORD, J. D., NITSCH, K. E., & FRANKS, J. J. Schooling and the facilitation of knowing. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, New Jersey: L. Erlbaum Associates, 1976.
- CRAIK, F. I. M., & KIRSNER, K. The effect of a speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, 1974, **26**, 274-284.
- CRAIK, F. I. M., & LOCKHART, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 671-684.
- CRAIK, F. I. M., & TULVING, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, **104**, 268-294.
- GARNER, W. R. *The processing of information and structure*. Potomac, Maryland: L. Erlbaum Associates, 1974.
- GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychological Review*, 1940, **28**, 93-115.
- GOLDSTEIN, A. G., & CHANCE, J. E. Visual recognition memory for complex configurations. *Perception and Psychophysics*, 1970, **9**, 237-241.
- HYDE, T. S., & JENKINS, J. J. Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology*, 1969, **82**, 472-481.
- JENKINS, J. J. Can we have a theory of meaningful memory? In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium*. Hillsdale, New Jersey: L. Erlbaum Associates, 1974.

- KIRSNER, K. An analysis of the visual component in recognition memory for verbal stimuli. *Memory and Cognition*, 1973, **1**, 449-453.
- KLEIN, K., & SALTZ, E. Specifying the mechanisms in a levels-of-processing approach to memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 671-679.
- KOLERS, P. A. Specificity of operations in sentence recognition. *Cognitive Psychology*, 1975a, **7**, 289-306.
- KOLERS, P. A. Memorial consequences of automatized encoding. *Journal of Experimental Psychology: Human Learning and Motivation*, 1975b, **1**, 689-701.
- KOLERS, P. A., & OSTRY, D. Time course of loss of information regarding pattern-analysing operations. *Journal of Verbal Learning and Verbal Behavior*, 1974, **13**, 599-612.
- LOCKHART, R. S., CRAIK, F. I. M., & JACOBY, L. L. Depth of processing, recognition and recall. In J. BROWN (Ed.), *Recognition and recall*. London: Wiley, 1975.
- MOSCOVITCH, M., & CRAIK, F. I. M. Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15**, 447-458.
- POSTMAN, L. Verbal learning and memory. *Annual Review of Psychology*, 1975, **26**, 291-335.
- SALTZ, E. *The cognitive bases of human learning*. Homewood, Illinois: The Dorsey Press, 1971.
- SCHULMAN, A. I. Memory for words recently classified. *Memory and Cognition*, 1974, **2**, 47-52.
- SEAMON, J. G., & MURRAY, P. Depth of processing in recall and recognition memory: Differential effects of stimulus meaningfulness and serial position. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 680-687.
- STEIN, B. S. The effects of cue-target uniqueness on cued-recall performance. *Memory and Cognition*, in press.
- TILL, R. E., & JENKINS, J. J. The effects of cued orienting tasks on the free recall of words. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 489-498.
- TULVING, E., & THOMSON, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, **80**, 352-373.
- TVERSKY, B. Encoding processes in recognition and recall. *Cognitive Psychology*, 1973, **5**, 275-287.
- WALSH, D. A., & JENKINS, J. J. Effects of orienting tasks on free recall in incidental learning: "Difficulty," "effort," and "process" explanations. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 481-488.

REFERENCE NOTE

1. HANNIGAN, M. L. *The effects of frameworks on sentence perception and memory*. Unpublished doctoral dissertation, Vanderbilt University, 1976.

(Received February 23, 1977)